RESEARCH ARTICLE

Double-blind validation of alternative wild bee identification techniques: DNA metabarcoding and *in vivo* determination in the field

Fernanda Herrera-Mesías[1,2]*, Christopher Bause[1]*, Sophie Ogan[3], Hannah Burger[4], Manfred Ayasse[4], Alexander M. Weigand[2], Thomas Eltz[1]

1 *Department of Animal Ecology, Evolution and Biodiversity, Ruhr-Universität, Bochum, Germany* **2** *Musée national d'histoire naturelle de Luxembourg, Luxembourg* **3** *Department of Biogeography, Trier University, Trier, Germany* **4** *Institute of Evolutionary Ecology and Conservation Genomics, Ulm University, Ulm, Germany*

Corresponding author: Thomas Eltz (thomas.eltz@rub.de)

Citation: Herrera-Mesías F, Bause C, Ogan S, Burger H, Ayasse M, Weigand AM, Eltz T (2022) Double-blind validation of alternative wild bee identification techniques: DNA metabarcoding and *in vivo* determination in the field. Journal of Hymenoptera Research 93: 189–214. https://doi.org/10.3897/jhr.93.86723

Abstract

Over the past few decades, several investigations around the globe have reported alarming declines in the abundance and diversity of bee species. The success of effective conservation strategies targeting these important pollinators relies heavily on accurate biodiversity assessments. The shortage of taxonomic experts and the escalation of the ongoing biodiversity crisis call for the development of alternative identification tools to implement efficient monitoring programs. The validation of such techniques is crucial to ensure that they provide results comparable to those of traditional morphotaxonomy. Here we performed two double-blind experiments to evaluate the accuracy of a pair of new techniques used for wild bee identification: DNA metabarcoding and *in vivo* identification in the field. The methods were tested on sets of wild bees from Germany and their results compared against evaluations done by panels of bee experts using traditional morphotaxonomy. On average the congruency of species identification between metabarcoding and morphotaxonomy was 88.98% across samples (N = 10), while *in vivo* identification and morphotaxonomy were 91.81% congruent (N = 7) for bees considered feasible for *in vivo* identification in the field. Traditional morphotaxonomy showed similar congruencies when compared to itself: 93.65% in the metabarcoding study and 92.96% in the *in vivo* study. Overall, these results support both new methods as viable alternatives to traditional microscopy-based assessment, with neither method being error-free. Metabarcoding provides a suitable option to analyze large numbers of specimens in the absence of highly

---

* These authors contributed equally to the article.

trained taxonomic experts, while *in vivo* identification is recommended for repeated long-term monitoring, and when working in areas where the sampling of individuals could threaten local populations of endangered wild bee species. Further research is still needed to explore the potential of both techniques for conservation management and wildlife monitoring, as well as to overcome their current limitations as taxonomic tools.

Introduction

Wild bees (Hymenoptera, Anthophila) are insect pollinators that are both ecologically important and of remarkable economic interest (Brown and Paxton 2009; Papanikolaou et al. 2017). As such, they are a key component of the global biodiversity, providing ecosystem services to wild flowering plants and commercially grown crops (Potts et al. 2010). Their services have a direct impact on food production. Not only do 75% of the world food crops benefit from insect-mediated pollination, mostly performed by bees, but it is estimated that about 42% of the leading crops grown for direct human consumption are pollinated by at least one wild bee species (Klein et al. 2007; Potts et al. 2010).

The recent decline of wild bees and other major insect groups in several regions of the world has become a matter of global concern among conservation biologists and the general public (Biesmeijer et al. 2006; Potts et al. 2015; Hallmann et al. 2017; Wagner 2020). The underlying causes for this decline are variable and still under investigation, but habitat loss and fragmentation, as well as agricultural pesticides and climate change, are mentioned as major drivers (Winfree et al. 2009; Hofmann et al. 2018; Meeus et al. 2018).

To preserve wild bee biodiversity, conservation initiatives adapted to the habitat requirements of local bee communities must be implemented (Müller et al. 2006; Brown and Paxton 2009; Henry and Rodet 2018; Ganser et al. 2021). The success of these conservation efforts relies heavily on accurate taxonomic information. Detailed knowledge regarding local species composition is key to selecting adequate strategies for habitat management and preservation (Ji et al. 2013).

Despite its importance, reliable taxonomic information is rather incomplete in several regions of the world. Even in Central Europe, the population trend of most wild bee species remains unknown (Potts et al. 2010; Gueuning et al. 2019). An estimate of 1,101 species in Europe (56.7% of the total) are classified as "data deficient" according to the European Red List of Bees, indicating a lack of scientific information to assess extinction risk (Nieto et al. 2014). Changes in regional bee fauna are poorly understood due to the lack of long-term insect monitoring programs, but there is evidence of local decline in species richness and community composition shifts (Hallmann et

al. 2017; Hofmann et al. 2018; Rollin et al. 2020). In Germany, about half of the occurring 550+ species of wild bees are categorized as threatened, based on Red List evaluations (Westrich et al. 2011; Schneider 2018; Vereecken 2018; Westrich 2019; Hofmann and Renner 2020). Conservation projects aiming to protect local wild bee populations must first retrieve accurate taxonomic information regarding which species are present in the area of interest, applying reliable taxonomic tools.

It is a common procedure in wild bee monitoring to collect adult specimens in the field via active methods such as targeted sweep netting, or passive sampling using devices like pan traps or vane traps (Roulston et al. 2007; Westphal et al. 2008; Falk 2016; Prendergast et al. 2020). The collected specimens are pinned, labeled and prepared for identification using a stereo microscope and morphological keys (Westrich 2019). Identification of pinned specimens based on morphological traits ("*PIN*") is the current gold standard for bee inventories. However, there are situations when *PIN* has shortcomings, especially (1) in the context of multi-replicate inventories over large spatial scales that are prone to exceed the available funds for or the capacity of classical morphological identification (Yu et al. 2012; Lebuhn et al. 2013; Gueuning et al. 2019), (2) for reduced-impact bee monitoring in areas where collecting/killing all individuals would risk exterminating local populations of rare species (Gezon et al. 2015) and (3) in cases of challenging morpho-identification (i.e. cryptic species complexes) (Schmidt et al. 2015). While these three challenges arise from quite different aspects of *PIN*, they are all serious concerns that are intensively discussed among wild bee experts (VDI-Richtlinie 4340-1, 2021).

The accuracy of *PIN* relies strongly on the experience of the taxonomist because it can be extraordinarily complex, as diagnostic traits can vary substantially between regions, localities, or even within local populations. Traits, especially coloration and vestiture, can even vary for a given individual bee over the flight season (Falk 2016). While in some taxonomic groups traits are well differentiated, in others the character states overlap and identifications require evaluation of combinations of traits, making unambiguous classification challenging even for trained experts (Michener 2000). In some bee genera reliable identification requires access to an established reference collection, a resource that is not always available (Gibbs et al. 2013). Due to these challenges, reliable *PIN* of large numbers of specimens is costly and may be precluded by the limited availability of trained taxonomic experts (Hopkins and Freckleton 2002; Engel et al. 2021).

DNA-based monitoring methods and molecular identification pipelines have great potential to assist *PIN* in wild bee inventorying (Gueuning et al. 2019). DNA metabarcoding is a molecular identification technique that relies on PCR primers for mass-amplification of taxonomically informative gene regions from bulk samples, combining high throughput sequencing (HTS) and parallel DNA-based species identification using bioinformatic tools to compile taxonomic lists up to species level (Ji et al. 2013; Brandon-Mong et al. 2015). It represents an upscaling to traditional Sanger sequencing DNA barcodes, as it allows the analysis of thousands of specimens simultaneously, assessing biodiversity rapidly and cost-efficiently (Yu et al. 2012), regardless of the life

stage of the specimens or their sex. Also, it provides an objective way to discriminate cryptic sibling species (Elbrecht and Leese 2015).

Despite their advantages, metabarcoding approaches are not free of technical limitations and flaws. Several investigations have reported that it is generally not possible to retrieve taxon abundance data because final read numbers are heavily affected by species amplification efficiency (i.e. primer bias; Zhou et al. 2013; Elbrecht and Leese 2015; Gueuning et al. 2019; Piñol et al. 2019). Moreover, results can be affected by other error sources leading to false positives (e.g. environmental contaminations), false negatives (e.g. gaps in the barcode reference libraries and significant biomass differences of specimens) or to discrepancies with traditional taxonomic outcomes (hybridization and shared barcodes among more recently diverged species) (Sheffield et al. 2009; Clarke et al. 2014; Schmidt et al. 2015; Elbrecht and Leese 2015; Weigand et al. 2019; Zinger et al. 2019). Therefore, the performance of metabarcoding approaches targeting wild bees must be cross-validated to ensure that robust data is produced for its use in conservation biology.

In the present study we test the accuracy of a customized metabarcoding pipeline ("*DNA*") incorporating a voucher-saving work-flow targeting Central European wild bees (Herrera-Mesías et al. submitted).

Both *PIN* and *DNA* metabarcoding of bulk samples, are invasive techniques in the sense that they remove specimens from the population, thereby reducing local population size and potentially endangering local population survival. Only very few studies are dealing with effects of such lethal sampling methods on population development. Even though Gezon et al. (2015) found no evidence for harmful effects of repeated, lethal sampling of bees, this might still be an important factor for species with very small population size or, in case of traps being used, for species that are particularly attracted to the type of trap (e.g. colored vane traps, Gibbs et al. 2017). To minimize such potential effects, Schindler et al. (2013) proposed a set of low-impact monitoring rules, which has been further developed in the BienABest project (www.bienabest.de) aiming to safeguard the ecosystem service of pollination and to enhance wild bee diversity in agricultural landscapes. The method, which has already been used in bee surveys within BienABest (Neumüller et al. 2020, 2021), has been elaborated in detail by VDI-Richtlinie 4340-1 (2021). It relies on identifying the majority of encountered bee specimens alive in the field, either by on-sight observation (e.g., on flowers) or by capture, brief confinement and immediate release following identification. The method is abbreviated as *IVI* in the present article (for *in vivo* identification). *IVI* is aimed to reduce negative impacts on the entire bee community, but in particular on species that are vulnerable and can be recognized with reasonable certainty directly in the field. It is also thought to improve data quality for long-term bee monitoring by reducing the effects of monitoring itself on the results, i.e. in case of repeated sampling in the same restricted bee habitats. Even more than *PIN*, *IVI* relies on trained and experienced bee experts that are capable of identifying many bee species directly in the field, without microscope and without consulting a reference collection, solely assisted by hand-net, observation jar, magnifying glass and identification keys.

Thus, for this study, double-blind experiments were performed to evaluate the accuracy of two alternative taxonomic identification techniques used on wild bees,

DNA metabarcoding of bulk samples ("*DNA*"), and *in vivo* identification ("*IVI*"). We compared the output of both methods against the evaluation of a panel of wild bee experts to determine similarities and discrepancies between the new approaches and traditional morphotaxonomy based on dry-pinned specimens ("*PIN*").

Materials and methods

DNA - Wild bee sampling and double-blind approach

To evaluate the metabarcoding pipeline described in Herrera-Mesías et al. (submitted) a total of 230 wild bee specimens were used. The samples were collected by S.O. and a field assistant using hand nets during 10 sampling events from 27 April to 22 July in 2020 in 7 different sites distributed across the Federal State of Rhineland-Palatinate (Germany). The netted bees were killed with ethyl acetate and immediately stored under cool conditions. From the end of the field day until the pinning of the individuals, all samples were stored frozen to prevent possible degradation of the DNA. Bees were pinned (males with genitals pulled out) and labeled by the end of the field season. For DNA extraction one complete midleg of each individual was removed using fire-sterilized tweezers and transferred to 2 mL Eppendorf tubes. After processing the bees of a sampling event, all surfaces and tools, i.e., tweezers, were sterilized to exclude cross contamination. The legs were pooled per sampling event, the pooled samples labeled with integers 1 through 10 by S.O. and shipped to the Zoology Department of the Musée national d'histoire naturelle Luxembourg, where further molecular analysis ("*DNA*") was conducted by F.H-M. and A.W. without specific knowledge of sites or specimens.

The pinned voucher specimens were shipped to two internationally recognized wild bee experts, both with over 15 years of experience in wild bee faunistics and taxonomy, who were asked to identify them to species level ("*PIN*"). Both experts consented the use of their identifications for the double-blind evaluation of the metabarcoding approach. During the laboratory analysis, the team processing the pooled leg samples had no access to the voucher specimens nor any of their metadata information or the evaluations done by the experts. The wild bee experts never met each other, and their taxon lists were handled by a third party (T.E.) until the DNA pipeline output was completed. The voucher specimens are deposited in the MNHNL invertebrate dry collection for long-term storage and curation (MNHNL127130-127359).

DNA - Metabarcoding pipeline

For the metabarcoding pipeline, a two-step PCR protocol using fusion primers based on Elbrecht and Steinke (2019) was used. The tags used for the second PCR are described in Elbrecht and Leese (2017). The laboratory protocols of Weigand and Herrera-Mesías (2020) were used for DNA extraction, as well as for the first PCR. For the

second PCR, 1 µl of the amplicon (without cleanup) was used as a template and the reaction volume was modified to a final volume of 50 µl. Both PCRs were run on an Eppendorf Mastercycler nexus eco Thermocycler using programs based on Elbrecht and Steinke (2019) and described in Herrera-Mesías et al. (submitted).

To increase the data robustness and the probability of detecting low biomass specimens, a PCR replicate strategy was followed. Two replicates of each sample plus one positive control (i.e. a mock community of known wild bee community composition) were included in the final setup. The success of both PCR replicates was verified by electrophoresis and their amplicons were purified with a NucleoSpin Gel and PCR Clean-up kit (Macherey-NagelTM). The DNA concentrations of the purified products were measured and equimolarly pooled into the final library (27.42 µl, 48.47 ng/µl). The clean library was sequenced on one lane of an Illumina MiSeq System (2x250 bp) at the Luxembourg Centre for Systems Biomedicine (Belval, Luxembourg).

The resulting DNA metabarcoding sequence data was processed using the JAMP R package (https://github.com/VascoElbrecht/JAMP), with the settings and supplementary tools described in Herrera-Mesías et al. (submitted). Taxonomic sorting was performed by comparing the resulting OTU fasta files against sequences stored in the Barcode of Life Data system (BOLD; Ratnasingham and Hebert 2007) using BOLDigger (Buchner and Leese 2020). As the team performing the bioinformatic analysis was blind to any metadata regarding the potential species composition of the samples, the default thresholds of BOLDigger were considered to find the best fitting hit for OTU taxonomic identification: at least a match of 85% for identification to the level of order, 90% to the level of family, 95% to the level of genus and 98% to the level of species.

The resulting data were pruned using TaxonTableTools (Macher et al. 2021) to remove all non-Hymenoptera OTUs, as well as Hymenoptera OTUs present in only one PCR replicate. Finally, the taxon name assignation of the filtered data was manually reviewed and partly modified from the original BOLD output by A.W. (blind to *PIN* results) to comply with current taxonomic nomenclature, thus creating a curated taxon list (Suppl. material 1). Only Hymenoptera OTUs present in both replicates with read numbers above 0.01% of abundance for each replicate and identified to species level were included in the final curated table. If a species was represented by multiple OTUs in the dataset, the results were collapsed into a single species entry.

To maintain double-blindness between *DNA* and *PIN*, the curated table was sent to T.E. who cross-tabulated identification results for each sample for a first comparison. Only then were the results made available to the rest of the team for numerical analysis. To allow comparison among the output of both approaches, the curated taxon list was transformed into a presence/absence table and combined with the results of the morphological approach.

IVI - Wild bee sampling and double-blind approach

To test the accuracy of *in vivo* determination of wild bees in the field, one of the authors (C.B.) accompanied bee monitorers during wild bee surveys within the "Bien-ABest" project. Surveys took place from April to September 2020 at nine different sites

throughout Germany and were conducted by a total of seven trained bee monitorers, whose experience in bee faunistics and taxonomy varied from some to many years. The monitorers used a reduced-impact monitoring method that includes *in vivo* identification (*IVI*) of encountered wild bees along variable transect walks (Neumüller et al. 2020, 2021, VDI-Richtlinie 4340-1 2021). Bees were either identified by the monitorer "on sight" when no closer scrutiny was deemed necessary, or were captured and identified with the help of an observation jar and a magnifier (ID method "capture"). Bees that could still not be identified *in vivo* were killed for later identification under the microscope. Overall, a total of 552 bee individuals were encountered by the seven monitorers during the surveys, of which 56 individuals (10.14%) were deemed impossible to identify in the field. The remaining 496 individuals were identified alive "on sight" or following "capture" by the monitorer. Of those, 210 individuals (42.34%) were consecutively collected by C.B. and stored in pre-labeled vials for later validations (see below). The remaining 287 individuals either could not be captured or were excluded from the evaluation because they represented species that had already been identified three times by an individual monitorer. This exclusion rule treated sexes separately, i.e., the maximum number of *IVI* individuals evaluated per species and monitorer was six (three females and three males).

The 210 bees to be included in the laboratory evaluation of *IVI* were killed with ethyl acetate or by freezing, and pinned by C.B. Furthermore, genitalia of male specimens were extracted and fixed outside the metasoma if required for species identification. The pinned specimens were re-labeled with a unique number code to omit information about date, locality or any other detail that would violate the anonymity of the monitorers.

The pinned bees were first identified by one internationally recognized wild bee expert with many years of experience in bee faunistics, morphotaxonomy and systematics (EXP data set) who worked under the knowledge that the identifications would later be used for *IVI* evaluation. Consecutively the specimens were sent to four other recognized wild bee experts, with several to many years of experience in bee morphotaxonomy, for independent identification (*PIN*). These experts were paid at rates typical for freelance work and were also aware that their work was part of a scientific investigation. To reconcile all discrepancies of identifications between the EXP data set and *PIN*, these were consecutively discussed in detail with the respective *PIN*-experts. Based on these discussions, and taking into account COI barcodes of two critical bee individuals (see Suppl. material 2 for laboratory protocol), a consensus list (CON data set) was established that represents our most objective assignment of true species affiliation. The voucher specimens are deposited in the collection of the Department for Evolutionary Ecology and Conservation Genomics at the University of Ulm.

For data analysis, the whole data set of wild bee IDs was divided into seven bee sets, each representing the identifications made by one individual monitorer, enabling us to analyze discrepancies between *IVI* and *PIN* across monitorers, and to contrast them with the discrepancy among *PIN* identifications for the same sets of bees. Additionally, a comparison to the consensus list showed the percentage of correctly identified bees per *IVI* and *PIN* expert.

Similarity analysis (*DNA* and *IVI*)

To further analyze the congruency and discrepancy of identification within and among *DNA* and *PIN*, and within and among *IVI* and *PIN*, we calculated Bray-Curtis similarities based on presence/absence taxon tables (*DNA* evaluation) or quantitative taxon tables (*IVI* evaluation) using the PRIMER-E software (version 6.1.6; Clarke and Gorley 2001), which was also used to plot dendrograms (hierarchical cluster analysis, complete linkage) based on the calculated similarity matrices.

Results

DNA - Evaluation of metabarcoding

After trimming and quality filtering, 2,874,629 high quality reads from the original 4,395,456 read pairs were retained (Short Read Archive bioproject: PRJNA876388). About 67.8% of the 1,447,238 original unassigned reads corresponded to PhiX. A total of 17.27% of the original 278 OTUs detected in the dataset were discarded after filtering based on a 0.01% read abundance threshold, remaining 230 OTUs for further analysis. 480 chimeras were discarded as well during clustering. After comparison against the BOLD systems database and replicate consistency analysis with Taxon Table Tools,146 OTU consistently found across replicates were preliminary identified as Hymenoptera taxa to various levels of taxonomic resolution (Suppl. material 3). After filtering, data merging and curation, 91 distinct taxonomic units representing detected wild bee species and species groups were included in the final curated table comparing *DNA* with *PIN* (Suppl. material 4).

The number of taxonomic units detected by *DNA* in individual samples varied between 11 and 22. All the species intentionally pooled in the mock community sample (positive control) were detected. From the ten samples considered in the analysis, only one (S2) presented a perfect congruence between the metabarcoding results ("*DNA*") and the evaluations of both taxonomic wild bee experts ("*PIN*1" and "*PIN*6"), based on the values of the Bray-Curtis index and the visual analysis of the dendrogram (Fig. 1). Two more DNA-based species lists were identical to the *PIN*1 expert results (S7, S9), none to *PIN*6. In six of the remaining samples, the DNA pipeline outcome was grouped closer to *PIN*1 on a terminal branch, with higher similarity than the resulting one from the comparison of the results of both experts. In three samples (S1, S3 and S8), the results of both *PIN* experts were more similar with each other than with the results of the DNA pipeline. When results are considered within the same sample, the Bray-Curtis similarity was 80% or higher among all three methods, with the lowest similarity observed between the pipeline and both experts in S3.

Across samples, the average congruency between the *DNA* and *PIN* ("*PIN*av") was 88.98% (Table 1). The mean congruency within *PIN* was slightly higher (93.65%). When *PIN* identifications were considered separately, the results of *DNA* were in better agreement with the evaluation of the first expert ("*PIN*1") than with the second one
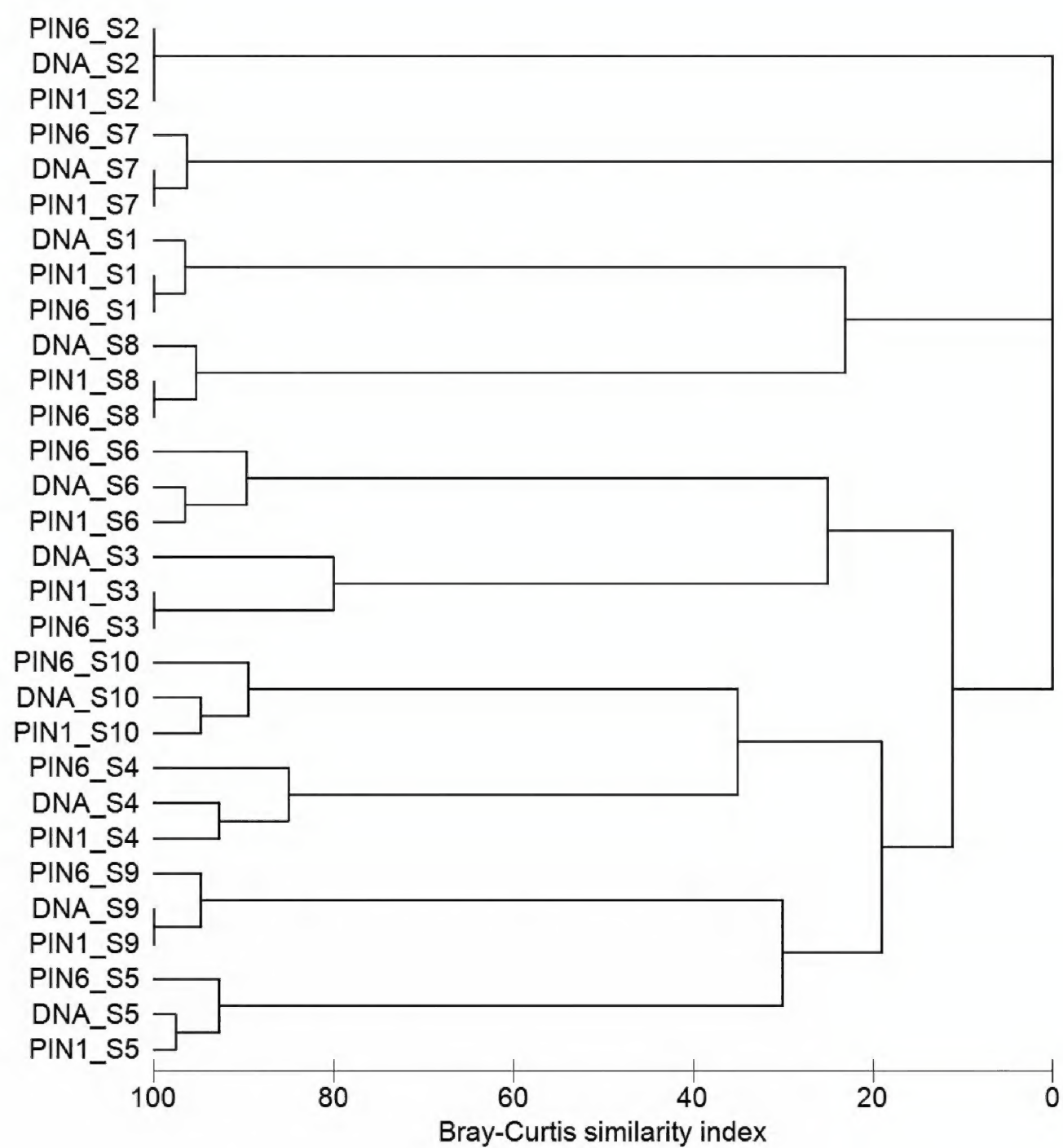
Figure 1. Dendrogram based on Bray-Curtis index to illustrate congruency of wild bee identification among the metabarcoding pipeline (*DNA*) and two *PIN* experts. X-axis shows index values expressed as percentage Bray-Curtis similarity.

Table 1. Percentage congruency of taxon lists resulting from *DNA* and *PIN* across samples. DNA x PIN1, DNA x PIN6 and PIN1 x PIN6: Percentages are calculated based only on the wild bee taxa detected by the methods considered in each pairwise comparison. DNA x PINav: Average of pairwise congruency between *DNA* and both *PIN* experts. Mean congruency across samples and standard deviations (SD) are also given. N = number of bee individuals in each set.

|	Bee Set (=Sample)											
Comparison	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean congruency (%)	SD
	N=25	N=14	N=21	N=28	N=30	N=27	N=17	N=14	N=27	N=27		
DNA x PIN1	93.33	100.00	66.67	90.48	95.24	93.33	100.00	90.91	100.00	90.00	92.00	9.75
DNA x PIN6	93.33	100.00	66.67	77.27	86.36	81.25	92.86	90.91	90.00	80.95	85.96	9.63
DNA x PINav	93.33	100.00	66.67	83.87	90.80	87.29	96.43	90.91	95.00	85.48	88.98	9.30
PIN1 x PIN6	100.00	100.00	100.00	85.71	90.48	87.50	92.86	100.00	90.00	90.00	93.65	5.77
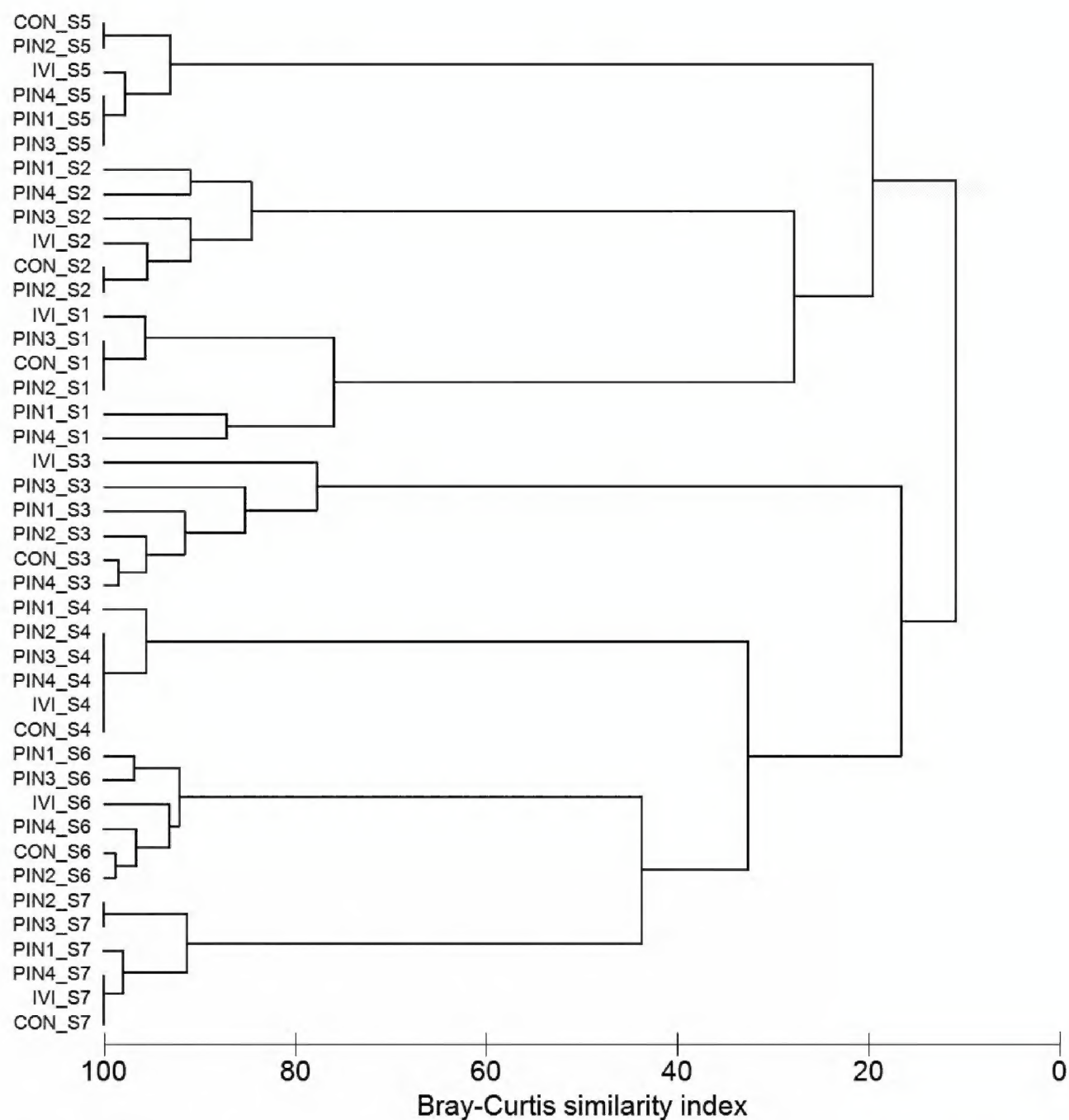
Figure 2. Dendrogram based on Bray-Curtis similarity index to illustrate congruency of wild bee identification among *IVI* and *PIN* experts. X-axis shows index values expressed as percentage Bray-Curtis similarity. *PIN* experts 1 to 4 are the same among the seven bee sets, whereas the *IVI* expert is different for each bee set.

Table 2. Percentage congruency of taxon lists resulting from *IVI* and *PIN* in comparison to each other and a curated consensus list (CON) for each of the seven bee sets. IVI x PIN: Average of pairwise congruencies between one *IVI* expert and the four *PIN* experts. PIN x PIN: Average of pairwise congruencies between the four *PIN* experts. IVI x CON: Congruency between one *IVI* expert and the consensus list. PIN x CON: Average of congruencies between each of the four PIN experts and the consensus list. Grand means and standard deviations (SD) across bee sets are also given. N = number of bee individuals in each set.

	Bee Set (=Sample)								
Comparison	S1	S2	S3	S4	S5	S6	S7	Mean congruency (%)	SD
	N=21	N=29	N=39	N=19	N=24	N=46	N=30		
IVI x PIN	86.90	92.24	80.77	98.68	94.79	93.48	95.83	91.81	5.62
PIN x PIN	85.71	91.38	90.60	97.37	97.92	93.84	93.89	92.96	3.90
IVI x CON	95.24	96.55	84.62	100.00	91.67	97.83	100.00	95.13	5.06
PIN x CON	90.48	93.97	94.87	98.68	96.88	95.65	95.83	95.19	2.38

("*PIN6*") or with their average outcome. The highest disagreement between *DNA* and *PIN* was observed in S3, where *DNA* detected five additional species and missed one identified by both *PIN* experts, reaching a congruency of only 66.67%.

IVI – Evaluation of *in vivo* identification

The total sample size of evaluated bees was reduced from originally 210 bees to 208 bees due to critical damage in two specimens caused by repeated shipping. The number of identified bees per monitorer/bee set varied from 19 to 46 bee individuals. Fig. 2 shows the similarity of species identifications by *IVI*, *PIN* and the CON data set based on the Bray-Curtis similarity index. The greatest congruency was found in bee set 4, in which the monitorer (*IVI4*) and three out of four *PIN* experts as well as the CON data set produced a perfectly identical taxon list. In bee set S2, S3 and S6 identification results differed at least slightly among the consulted *IVI* and *PIN* experts. The largest discrepancies were found in bee sets S1 and S3 (86.90% and 80.77%, respectively).

Averaged across bee sets, there was a taxon list congruency between *IVI* and *PIN* of 91.81%. *PIN* results among themselves showed an average taxon list congruency of 92.96% (Table 2). Overall, and in comparison with the CON dataset, the average percentage of correctly identified bee individuals was 95.13% for *IVI* and 95.19% for *PIN* (see Table 2). Apparent misidentifications of *IVI* and *PIN* experts appeared mainly within bee genera *Andrena*, *Bombus*, *Halictus* and *Megachile*. In addition, some bee individuals of *Lasioglossum* spp. were misidentified by *PIN* experts (Suppl. material 5).

Discussion

The performed double-blind validations demonstrated that error rates of the evaluated novel methods were of a similar (low) order of magnitude as compared to traditional morphotaxonomy, suggesting they represent valid alternatives for wild bee monitoring. In addition, we found that neither of the methods, traditional pinning, *in vivo* identification or DNA metabarcoding, were error free. In the following we shed light on the types of errors that occurred and discuss strengths and weaknesses of the respective methods. To our knowledge, this is the first double-blind study to evaluate per-sample accuracy of wild bee identification within and across methods. Even if previous studies have compared the congruency of diverse identification techniques used in wild monitoring against traditional morphotaxonomic outcomes (Tang et al. 2015; Gueuning et al. 2019), this is the first experiment to date that has been explicitly designed to control the bias resulting from the exchange of preliminary taxonomic information among the different participants, thus to ensure that the results are based purely on the detection capacity of each identification technique.

Evaluation of DNA metabarcoding in comparison with morphotaxonomy

The overall congruency found between the metabarcoding pipeline (*DNA*) and morphological identification results (*PIN*) on a per-sample basis analysis (88.98% mean congruency) agrees well with previous findings reported by Gueuning et al. (2019). In their study, based on a multi locality setting in Switzerland, over 90% of the traditionally identified morphospecies were also detected by DNA metabarcoding.

Despite the high overall similarity of the results obtained by *DNA* and *PIN* in our study, 26 cases of disagreement were present (Suppl. material 4), which are worth further discussion: In 12 cases, the molecular results support the assessment of one morphotaxonomic expert against the other, resolving conflicting morphological evaluations. Incongruence between *DNA* and both *PIN* experts can partially be explained by unclear species delimitation. There is a historical controversy regarding whether *Andrena ovatula* and *Andrena albofasciata* should be consider as one or two species (Westrich et al. 2011; Schmidt et al. 2015; Praz et al. 2022). In our study, the metabarcoding pipeline supported the presence of *A. albofasciata* against *A. ovatula* in S4 and S10, in opposition to the morphological analysis, but was in agreement with both *PIN* experts regarding detecting only *A. ovatula* in S5. As *DNA* recognized these taxa as two separate OTUs in our dataset based on a 97% genetic similarity threshold, this suggests the presence of a second species, potentially overlooked by *PIN*, within what has been traditionally considered *Andrena ovatula sensu lato*. These results are in agreement with recent analyses that have resolved the controversy by consistently demonstrating the existence of two distinct species within the complex, *A. ovatula* and *A. afzeliella* (Kirby, 1802) (=*A. albofasciata*), based on molecular, morphological and ecological evidence (Praz et al. 2022). Therefore, the nomenclature of DNA barcodes currently available in BOLD should be updated accordingly to match this new taxonomic consensus, further improving the detection capacity of molecular approaches.

Further research on cryptic diversity following a similar approach would contribute to reach final conclusions regarding the status of similarly challenging species complexes, such as the *Halictus simplex*-complex. Although our dataset pooled species within this complex into one entity for the overall comparisons, *DNA* was able to precisely identify *H. langobardicus* regardless of the sex of the individual, whereas *PIN* was only able to assign a species-level annotation to males (Suppl. material 4).

Given that the genetic results of controversial species complexes involve an additional level of analysis (Schmidt et al. 2015), a sufficient number of validated DNA reference barcodes should be a pre-requirement to perform metabarcoding on taxonomically problematic sibling species. Whenever possible, barcodes from local specimens reliably identified by known taxonomic experts should be preferred as reference material, thus to reach accurate interpretations.

Another factor potentially affecting the congruency of metabarcoding results with morphological analysis is environmental contamination. For example, in seven cases the pipeline detected additional wild bee species to the ones reported by the taxonomic experts. Five false positive detections were found in S3 (*Bombus lapidarius*, *Bombus*

pascuorum, *Andrena cineraria*, *Chelostoma florisomne* and *Dasypoda hirtipes*), one in *S5* (*Halictus confusus*) and one in S8 (*Melecta luctuosa*). The additional species in S3 correspond to easily identifiable wild bees and three of them were completely absent in the whole wild bee set, which means that they cannot have been overlooked by *PIN*. Most likely, DNA traces from an outside source are likely responsible for these additional findings. Carry-over DNA from other specimens in the field, the sampling containers, or from specimen handling before DNA extraction represents a more likely explanation than cross-contamination in the laboratory as no other bees were being processed within the laboratory premises at the time of the double-blind experiment. The same situation may explain the presence of *H. confusus* in S5 and of *M. luctuosa* in S8. Tag-switching as an alternative explanation for the false positive results of species generally present in the overall data set seems unlikely, as tag combinations with high Levenshtein distances (=>3) were chosen to avoid the artificial generation of existing tag combinations given the sequencing platform used (Salipante et al. 2014; Elbrecht and Steinke 2019).

False positives and false negatives are known drawbacks affecting taxonomic assessment results originating from PCR-based high throughput sequencing techniques, potentially leading to taxonomic biases such as "biodiversity inflation" (Zhou et al. 2013; Tang et al. 2015; Gueuning et al. 2019). Identifying contaminants in wild bee metabarcoding datasets can be hard, because amplification bias may result in false positives, with read numbers equal or higher than the read numbers of true positives (Tang et al. 2015). Even if the false positive found in S8 had fewer reads than any true positive within the sample, their numbers were still over the defined threshold and similar to the read numbers of true positives found in other samples (see Suppl. material 3). Strategies that boost data robustness, such as increasing the number of PCR replicates of the same biological sample (Alberdi et al. 2018; Weigand and Macher 2018) or adjusting the value of filtering thresholds during bioinformatic pruning may be helpful to separate out potential false positives.

Finally, three false negatives were also found in the metabarcoding dataset (*Sphecodes gibbus* in S1, *Lasioglossum pauxillum* in S3 and *Melecta albifrons* in S6). In this case, insufficient sequencing depth seems a more likely explanation than obscurity due to primer bias, as all missing species show low primer-template mismatch with the selected primer pair (Herrera-Mesías et al. submitted). In the experiment, the sequencing run produced fewer overall read numbers than the ones reported by similar works (13.8 million reads in Gueuning, et al. (2019); 11.7 million in Herrera-Mesías et al. (submitted)). Compared to the 47,471 average reads per community of Gueuning et al. 2019, the average number of reads per sample replicate obtained in the double-blind experiment was almost three fold higher (134,340 reads after trimming and quality filtering). However, it was less than a third of the 460,074 average reads per replicate included in the final dataset of Herrera-Mesías et al. (submitted). Therefore, insufficient sequencing depth may have negatively affected specimens of low biomass represented by single individuals in certain sample mixtures. This seems to be the case for *L. pauxillum* in S3. The species presented 12 reads in the first replicate (threshold

of 14 reads) of the sample and 20 reads in the second (threshold of 14 reads), just barely below the 0.01% inclusion threshold (Suppl. material 3). Adjusting the pooling scheme of the library considering different criteria and additional metadata regarding the sample in question (i.e., final DNA concentration in relation to the number of specimens for each bulk sample, size sorting, etc.) may help to reduce the likelihood of false negatives.

Despite the lack of a perfect match with the expert evaluations, the results of the DNA metabarcoding pipeline are similar enough to be advised as a viable alternative to microscopy-based assessment, especially when considering its high congruency to the *PIN*1 results. Moreover, this approach offers several advantages for broad-scale assessments in the context of conservation biology projects, when large quantities of wild bees may be challenging and costly to identify (Lebuhn et al. 2013; Creedy et al. 2020). The number of specimens here analyzed could be increased 10-fold without substantially rising laboratory expenses, work effort, or compromising the quality of results. However, increasing the number of samples can also reduce the number of sequences per replicate, potentially increasing the risk of false negatives. Therefore, each analysis must consider the desired sequencing depth per sample as well as the performance of the platform selected to determine the maximum number of samples that can be pooled on the same run (Elbrecht and Steinke 2019).

Finally, DNA metabarcoding presents a crucial limitation for wild bee monitoring purposes, as it should only be used for qualitative assessment. An alternative molecular, cost-effective but specimen-based solution allowing qualitative results can be offered by high-throughput or next-generation sequencing DNA barcoding (Creedy et al. 2019; Gueuning et al 2019).

Evaluation of In-Vivo Identification

We found that *IVI* of bee individuals considered feasible for alive determination in the field by the monitorer led to similar rates of correct identification as *PIN*, i.e., 95% as judged post-hoc based on the curated consensus list (CON). This may seem surprising, because *IVI* took place in the field without a dissecting microscope. For a better understanding of the results, it is necessary to look more closely at the different error sources that led to incongruencies between the expert identifications.

First, biased expectations appeared to have caused misidentifications especially in *IVI*, where monitorers had knowledge of local bee communities from previous visits. This kind of mistake seems to have generated several cases of incorrect bumblebee identification. For example, in case of BBV86 and BBV98 (see Suppl. material 5) a similar but more noteworthy species was chosen instead of the abundant *Bombus lapidarius*. In another case a female *Megachile leachella* (BBV188) was confused with *Megachile pilidens*. Whereas *M. leachella* was not previously known to occur in the locality, the similar *M. pilidens* had been expected from previous encounters (pers. comm. of monitorer with C.B.).

In contrast, *PIN* appears to have been more susceptible to mistakes like misplaced entries in excel sheets or mix-up of specimens. Such errors were suggested by unlikely misidentifications as in BBV42, a worker bumblebee *Bombus lapidarius* that had been identified as *Halictus subauratus*, a bee that could not be more different. In addition, some errors arose from biases in the used identification keys or reference collections. For example, the popular (and generally very good) identification key for bumblebees by Mauss (1987) does not cover the full range of (corbicula hair) color variability of *Bombus humilis*. Its use by *PIN* experts was associated with repeated misidentification of *Bombus humilis* workers as *Bombus ruderarius*, for which reddish corbicula hair is a well known trait (BBV13, BBV14, BBV16, pers. comm. with C.B.). The alternative distinctive trait (shape of labrum bottom edge) given by the key was not considered by the experts and another evident characteristic of the specimens (bright facial hair; untypical for *B. ruderarius*) was neither explicitly treated by the key nor noticed by the experts. Biased reference collections appeared to have caused other errors in *PIN*. For example, the expert who incorrectly identified a female of *Megachile maritima* (BBV169) as *Megachile willughbiella* did so based on divergent reference material collected from populations outside of Germany (pers. comm. with C.B.). In discussions with C.B., some *PIN* experts stated their insufficient experience with species outside of their region of expertise as a possible source of error.

Due to the design of the study there might be a number of intrinsic biases that could have increased the accuracy of *IVI* relative to *PIN*. First, *IVI* experts had a free choice regarding which of the encountered bees they considered feasible for *IVI* (during evaluated monitorings approximately 90% of individuals were considered feasible for *IVI*, a rate that corresponds well with *IVI* rates during regular BienABest monitorings; BienABest project, unpublished results). Thus, they could directly influence the sample of bees/identifications that was being evaluated. In addition, *IVI* experts were very aware of being evaluated, and were constantly reminded of the fact by the presence of C.B. who collected their *IVI* bees. *PIN* experts, while also having been informed that their results will be used in a double-blind evaluation, did not work under close observation. This discrepancy in experienced scrutiny could have led to different likelihoods of careless mistakes.

The relationship between the amount of experience of the expert and the accuracy of identification results is less than clear. All experts included in this study (*IVI* and *PIN*, also for the *DNA* comparison) were recognized experts of bee morphotaxonomy with at least some years, but mostly many years, of experience. If there was a difference at all, the amount of experience was slightly higher and less variable among *PIN* than among *IVI* experts. The *IVI* monitorer considered least experienced did indeed deliver the least accurate identification result of only 84.6% in comparison to the consensus list. However, the respective bee set (S3) was also the one that had the lowest congruency among *PIN* experts (90.6%), suggesting that the set was difficult.

In general, *IVI* as conducted within the BienABest project yielded accurate identifications in nineteen out of twenty bees (95%). It needs to be emphasized that such accuracy can only be achieved by highly trained experts, a resource that is in short

supply (Drew 2011) and needs to be replenished by concerted efforts of universities, NGOs, national authorities and funding agencies. Probably, *IVI* will remain limited to a certain part of bee diversity that is feasible for *IVI*. Exactly how large this part is, is a matter of debate. According to a list ("Ampelliste") prepared by experts during their work on the VDI-Richtlinie 4340-1 (2021) just about 50% of females of German bee species can currently be identified alive. In the male sex the percentage is considered to be even lower (30%). It remains to be seen if this percentage can be increased in the future with the help of digital tools that allow scrutiny of additional taxonomical characters. Currently, such a tool is being developed within the BienABest project for identification of 300 bee species of Central Europe via smartphone app, which includes high quality pictures to guide reliable identification under field conditions.

There is a controversial debate on whether the use of *IVI* is in fact necessary and desirable for wild bee monitoring (Gezon et al. 2015). Generally, the effect of invasive sampling on insect populations, and bees in particular, has not been well studied (Packer and Darla-West 2021). We are aware of only one study that was dedicated to assess the effect on wild bee communities: Gezon et al. (2015) found no negative effects of several years of bi-weekly pan trapping and netting on bee communities in the Rocky Mountains (Colorado, USA). However, the study was conducted in large tracts of near natural habitat, and it is questionable whether the results can be transferred to the degraded and fragmented bee habitats in Central Europe (e.g., Steffan-Dewenter et al. 2006). It seems plausible to assume that repetitive removal of reproductive individuals can affect local populations of already endangered species, especially in solitary bees which are characterized by low reproductive rates and which often demonstrate a highly localized distribution (Westrich 2019). This is supported by at least one study that used colored vane traps and found conspicuous declines of attracted species in one locality (Gibbs et al. 2017). Depending on locality and monitoring design, *IVI* may be the way of erring on the safe side.

On reference specimens

IVI and most DNA metabarcoding approaches relying on bulk samples might have another disadvantage, as both strategies usually do not deposit extensive reference material. A reference collection for future comparison is often a legal requirement or at least important to judge about spatio-temporal patterns of individual species in times of changing taxonomies, e.g. within species complexes. Voucher specimens are also relevant in case upcoming taxonomic methods require biomaterial or morphometric data to address open taxonomic questions, or for educational purposes (Lister et al. 2011; Monfils et al. 2017; Kharouba et al. 2019), or to validate particularly noteworthy findings. Moreover, voucher specimens stored in local natural history collections represent an important resource for the construction of future taxonomic lists, including potentially overlooked findings relevant to the development of national conservation strategies (Herrera-Mesías and Weigand 2021). The most common referencing strategy of DNA metabarcoding approaches – if any – is the deposition of DNA vouchers. However, in cases of surprising results, DNA vouchers will make it difficult to further judge about the unexpected results.

In the metabarcoding setup here applied, DNA was extracted from individual legs while the rest of the voucher specimens were archived in the invertebrate collection of the MNHNL. Although this led to an increase in the hands-on-times and costs per sample, it preserves specimens for future conservation studies (Herrera-Mesías and Weigand 2021). Single specimen barcoding or HTS barcoding might also be helpful in the context of wild bee monitoring (Schmidt et al. 2015; Gueuning et al. 2019), especially when abundance data are desirable and total specimen numbers feasible to handle.

Regarding *IVI*, additional documentation could be provided by depositing high-quality images taken from live bees confined in observation jars, as is currently done by some experts. However, this requires appropriate equipment and imposes substantial additional effort during field work. Also, there currently exists no general depository for digital specimens of wild bees.

Conclusion

To our best knowledge, this is the first study to compare the accuracy of alternative taxonomic tools against morphology-based identifications using a double-blind approach. Both DNA metabarcoding and *in vivo* determination in the field presented high overall congruency of their identification results with a traditional microscopy-based assessment performed by morphotaxonomic experts. These results validate the use of these alternative assessment techniques in conservation projects targeting wild bees of Central Europe. The metabarcoding pipeline is recommended for the qualitative analysis of large samples in the absence of taxonomic experts, and for resolving morphotaxonomic problems. However, strategies that boost data robustness are highly advised to control the effect of potential environmental contaminations, false positives, and false negatives. Moreover, metabarcoding data should not be used on its own to estimate quantitative population parameters due to biases in PCR amplification. On the other hand, *in vivo* identification can be used for quantitative assessment. It is advised for long-term monitoring, especially in fragile ecosystems with vulnerable bee populations. It is susceptible to misidentification due to preconceptions and potentially constrained by the experience and availability of monitorers. By concept, *in vivo* identification results in no or fewer deposited reference specimens so that the detection of rare and particularly noteworthy species may be difficult to validate. Generally, all techniques rely heavily on the availability of reference materials such as barcode sequences, voucher specimens, or reference images. Further efforts are needed to address this issue, thus filling the gap of information needed to refine the detection capacity of alternative identification techniques.

Acknowledgements

References

Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution 9: 134–147. https://doi.org/10.1111/2041-210X.12849

Biesmeijer JC, Roberts SPM, Reemer M, Ohlemüller R, Edwards M, Peeters T, Schaffers AP, Potts SG, Kleukers R, Thomas CD, Settele J, Kunin WE (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313: 351–354. https://doi.org/10.1126/science.1127863

Brandon-Mong GJ, Gan HM, Sing KW, Lee PS, Lim PE, Wilson JJ (2015) DNA metabarcoding of insects and allies: An evaluation of primers and pipelines. Bulletin of Entomological Research 105: 717–727. https://doi.org/10.1017/S0007485315000681

Brown MJF, Paxton RJ (2009) The conservation of bees: A global perspective. Apidologie 40: 410–416. https://doi.org/10.1051/apido/2009019

Buchner D, Leese F (2020) BOLDigger - a Python package to identify and organise sequences with the Barcode of Life Data systems. Metabarcoding and Metagenomics 4: 19–21. https://doi.org/10.3897/mbmg.4.53535

Clarke KR, Gorley RN (2001) PRIMER v5: User manual/tutorial. primer e, Plymouth, UK.

Clarke LJ, Soubrier J, Weyrich LS, Cooper A (2014) Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. Molecular Ecology Resources 14. https://doi.org/10.1111/1755-0998.12265

Creedy TJ, Norman H, Tang CQ, Qing Chin K, Andujar C, Arribas P, O'Connor RS, Carvell C, Notton DG, Vogler AP (2020) A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. Molecular Ecology Resources 20: 40–53. https://doi.org/10.1111/1755-0998.13056

Drew LW (2011) Are We Losing the Science of Taxonomy? As need grows, numbers and training are failing to keep up. BioScience 61: 942–946. https://doi.org/10.1525/bio.2011.61.12.4

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS ONE 10. https://doi.org/10.1371/journal.pone.0130324

Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science 5. https://doi.org/10.3389/fenvs.2017.00011

Elbrecht V, Steinke D (2019) Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology 64: 380–387. https://doi.org/10.1111/fwb.13220

Engel MS, Ceríaco LMP, Daniel GM, Dellapé PM, Löbl I, Marinov M, Reis RE, Young MT, Dubois A, Agarwal I, Lehmann A. P, Alvarado M, Alvarez N, Andreone F, Araujo-Vieira K, Ascher JS, Baêta D, Baldo D, Bandeira SA, Barden P, Barrasso DA, Bendifallah L, Bockmann FA, Böhme W, Borkent A, Brandão CRF, Busack SD, Bybee SM, Channing A, Chatzimanolis S, Christenhusz MJM, Crisci J V, D'elía G, Da Costa LM, Davis SR, De Lucena CAS, Deuve T, Fernandes Elizalde S, Faivovich J, Farooq H, Ferguson AW, Gippoliti S, Gonçalves FMP, Gonzalez VH, Greenbaum E, Hinojosa-Díaz IA, Ineich I, Jiang J, Kahono S, Kury AB, Lucinda PHF, Lynch JD, Malécot V, Marques MP, Marris JWM, Mckellar RC, Mendes LF, Nihei SS, Nishikawa K, Ohler A, Orrico VGD, Ota H, Paiva J, Parrinha D, Pauwels OSG, Pereyra MO, Pestana LB, Pinheiro PDP, Prendini L, Prokop J, Rasmussen C, Rödel M-O, Rodrigues MT, Rodríguez SM, Salatnaya H, Sampaio Í, Sánchez-García A, Shebl MA, Santos BS, Solórzano-Kraemer MM, Sousa ACA, Stoev P, Teta P, Trape J-F, Dos Santos CV-D, Vasudevan K, Vink CJ, Vogel G, Wagner P, Wappler T, Ware JL, Wedmann S, Zacharie CK (2021) The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. Zoological Journal of the Linnean Society 193: 381–387. https://doi.org/10.1093/zoolinnean/zlab072

Falk S (2016) Field Guide to the Bees of Great Britain and Ireland. Bloomsbury Publishing (Bloomsbury): 1–432.

Ganser D, Albrecht M, Knop E (2021) Wildflower strips enhance wild bee reproductive success. Journal of Applied Ecology 58: 486–495. https://doi.org/10.1111/1365-2664.13778

Gezon ZJ, Wyman ES, Ascher JS, Inouye DW, Irwin RE (2015) The effect of repeated, lethal sampling on wild bee abundance and diversity. Methods in Ecology and Evolution 6: 1044–1054. https://doi.org/10.1111/2041-210X.12375

Gibbs J, Packer L, Dumesh S, Danforth BN (2013) Revision and reclassification of *Lasioglossum* (*Evylaeus*), *L.* (*Hemihalictus*) and *L.* (*Sphecodogastra*) in eastern North America (Hymenoptera: Apoidea: Halictidae). Zootaxa 3672: 1–117. https://doi.org/10.11646/zootaxa.3672.1.1

Gibbs J, Joshi NK, Wilson JK, Rothwell NL, Powers K, Haas M, Gut L, Biddinger DJ, Isaacs R (2017) Does passive sampling accurately reflect the bee (Apoidea: Anthophila) communities pollinating apple and sour cherry orchards? Environmental Entomology 46: 579–588. https://doi.org/10.1093/ee/nvx069

Gueuning M, Ganser D, Blaser S, Albrecht M, Knop E, Praz C, Frey JE (2019) Evaluating next-generation sequencing (NGS) methods for routine monitoring of wild bees: Metabarcoding, mitogenomics or NGS barcoding. Molecular Ecology Resources 19: 847–862. https://doi.org/10.1111/1755-0998.13013

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T, Goulson D, De Kroon H (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0185809

Henry M, Rodet G (2018) Controlling the impact of the managed honeybee on wild bees in protected areas. Scientific Reports 8: 1–10. https://doi.org/10.1038/s41598-018-27591-y

Herrera-Mesías F, Weigand AM (2021) Updates to the Checklist of the Wild bee Fauna of Luxembourg as Inferred from Revised Natural History Collection Data and Fieldwork. Biodiversity Data Journal 9: e64027. https://doi.org/10.3897/BDJ.9.E64027

Herrera-Mesías F, Kharrat Ep Jarboui I, Weigand AM (submitted) A metabarcoding framework for wild bee assessment in Luxembourg.

Hofmann MM, Renner SS (2020) Bee species decrease and increase between the 1990s and 2018 in large urban protected sites. Journal of Insect Conservation 24: 637–642. https://doi.org/10.1007/s10841-020-00238-y

Hofmann MM, Fleischmann A, Renner SS (2018) Changes in the bee fauna of a German botanical garden between 1997 and 2017, attributable to climate warming, not other parameters. Oecologia 187: 701–706. https://doi.org/10.1007/s00442-018-4110-x

Hopkins GW, Freckleton RP (2002) Declines in the numbers of amateur and professional taxonomists: Implications for conservation. Animal Conservation 5: 245–249. https://doi.org/10.1017/S1367943002002299

JAMP (2022) R package (metabarcoding pipeline). https://github.com/VascoElbrecht/JAMP

Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, Larsen TH, Hsu WW, Benedick S, Hamer KC, Wilcove DS, Bruce C, Wang X, Levi T, Lott M, Emerson BC, Yu DW (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecology Letters 16: 1245–1257. https://doi.org/10.1111/ele.12162

Kharouba HM, Lewthwaite JMM, Guralnick R, Kerr JT, Vellend M (2019) Using insect natural history collections to study global change impacts: Challenges and opportunities. Philosophical Transactions of the Royal Society B: Biological Sciences 374 pp. https://doi.org/10.1098/rstb.2017.0405

Klein AM, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T (2007) Importance of pollinators in changing landscapes for world crops. Proceedings of the Royal Society B: Biological Sciences 274: 303–313. https://doi.org/10.1098/rspb.2006.3721

Lebuhn G, Droege S, Connor EF, Gemmill-Herren B, Potts SG, Minckley RL, Griswold T, Jean R, Kula E, Roubik DW, Cane J, Wright KW, Frankie G, Parker F (2013) Detecting Insect Pollinator Declines on Regional and Global Scales. Conservation Biology 27: 113–120. https://doi.org/10.1111/j.1523-1739.2012.01962.x

Leese F, Bouchez A, Abarenkov K, Altermatt F, Borja Á, Bruce K, Ekrem T, Čiampor F, Čiamporová-Zaťovičová Z, Costa FO, Duarte S, Elbrecht V, Fontaneto D, Franc A, Geiger MF, Hering D, Kahlert M, Kalamujić Stroil B, Kelly M, Keskin E, Liska I, Mergen P, Meissner K, Pawlowski J, Penev L, Reyjol Y, Rotter A, Steinke D, van der Wal B, Vitecek S, Zimmermann J, Weigand AM (2018) Chapter Two - Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action. Advances in Ecological Research 58: 63–99. https://doi.org/10.1016/bs.aecr.2018.01.001

Lister, AM, and Climate Change Research Group (2011) Natural history collections as sources of long-term datasets. Trends in Ecology & Evolution 26: 153–154. https://doi.org/10.1016/j.tree.2010.12.009

Macher TH, Beermann AJ, Leese F (2021) TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. Molecular Ecology Resources 21: 1705–1714. https://doi.org/10.1111/1755-0998.13358

Mauss V (1987) Bestimmungsschlüssel für die Hummeln der Bundesrepublik Deutschland. Deutscher Jugendbund für Naturbeobachtung (Hamburg): 1–50.

Meeus I, Pisman M, Smagghe G, Piot N (2018) Interaction effects of different drivers of wild bee decline and their influence on host-pathogen dynamics. Current Opinion in Insect Science 26: 136–141. https://doi.org/10.1016/j.cois.2018.02.007

Michener CD (2000) The Bees of the World. The Johns Hopkins University Press (Baltimore and London): 1–877.

Monfils AK, Powers KE, Marshall CJ, Martine CT, Smith JF, Prather LA (2017) Natural History Collections: Teaching about Biodiversity Across Time, Space, and Digital Platforms. Southeastern Naturalist 16: 47–57. https://doi.org/10.1656/058.016.0sp1008

Müller A, Diener S, Schnyder S, Stutz K, Sedivy C, Dorn S (2006) Quantitative pollen requirements of solitary bees: Implications for bee conservation and the evolution of bee-flower relationships. Biological Conservation 130: 604–615. https://doi.org/10.1016/j.biocon.2006.01.023

Neumüller U, Burger H, Krausch S, Blüthgen N, Ayasse M (2020) Interactions of local habitat type, landscape composition and flower availability moderate wild bee communities. Landscape Ecology 35: 2209–2224. https://doi.org/10.1007/s10980-020-01096-4

Neumüller U, Burger H, Schwenninger HR, Hopfenmüller S, Krausch S, Weiß K, Ayasse M (2021) Prolonged blooming season of flower plantings increases wild bee abundance and richness in agricultural landscapes. Biodiversity and Conservation 30: 3003–3021. https://doi.org/10.1007/s10531-021-02233-4

Nieto A, Roberts SPM, Kemp J, Rasmont P, Kuhlmann M, García Criado M, Biesmeijer JC, Bogusch P, Dathe HH, De la Rúa P, De Meulemeester T, Dehon M, Dewulf A, Ortiz-Sánchez FJ, Lhomme P, Pauly A, Potts SG, Praz CQ, Window J, Michez D (2014) European Red List of Bees. Publication Office of the European Union (Luxembourg), 1–84. https://doi.org/10.2779/77003

Packer L, Darla-West G (2021) Bees: How and Why to Sample Them. In: Santos JC, Fernandes GW (Eds) Measuring Arthropod Biodiversity. Springer (Cham), 55–83. https://doi.org/10.1007/978-3-030-53226-0_3

Papanikolaou AD, Kuhn I, Frenzel M, Kuhlmann M, Poschlod P, Potts SG, Roberts SPM, Schweiger O (2017) Wild bee and floral diversity co-vary in response to the direct and indirect impacts of land use. Ecosphere 8(11): e02008. https://doi.org/10.1002/ecs2.2008

Piñol J, Senar MA, Symondson WOC (2019) The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. Molecular Ecology 28: 407–419. https://doi.org/10.1111/mec.14776

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: Trends, impacts and drivers. Trends in Ecology and Evolution 25: 345–353. https://doi.org/10.1016/j.tree.2010.01.007

Potts SG, Biesmeijer JC, Bommarco R, Breeze TD, Carvalheiro LG, Franzén M., González-Varo, Schweiger O (2015) Status and Trends of European Pollinators. Key findings of the STEP project. Pensoft Publishers (Sofia), 1- 72. www.step-project.de

Praz C, Genoud D, Vaucher K, Bénon D, Monks J, Wood TJ (2022) Unexpected levels of cryptic diversity in European bees of the genus *Andrena* subgenus *Taeniandrena* (Hymenoptera , Andrenidae): implications for conservation. Journal of Hymenoptera Research 428: 375–428. https://doi.org/10.3897/jhr.91.82761

Prendergast KS, Menz MHM, Dixon KW, Bateman PW (2020) The relative performance of sampling methods for native bees: an empirical test and review of the literature. Ecosphere 11(5): e03076. https://doi.org/10.1002/ecs2.3076

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System: Barcoding. Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rollin O, Vray S, Dendoncker N, Michez D, Dufrêne M, Rasmont P (2020) Drastic shifts in the Belgian bumblebee community over the last century. Biodiversity and Conservation 29: 2553–2573. https://doi.org/10.1007/s10531-020-01988-6

Roulston TH, Smith SA, Brewster AL (2007) Short communication: A comparison of pan trap and intensive net sampling techniques for documenting a bee (Hymenoptera: Apiformes) fauna. Journal of the Kansas Entomological Society 80: 179–181. https://doi.org/10.2317/0022-8567(2007)80[179:ACOPTA]2.0.CO;2

Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, Harkins TT, Cookson BT, Hoffman NG (2014) Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Applied and Environmental Microbiology 80: 7583–7591. https://doi.org/10.1128/AEM.02206-14

Sheffield CS, Hebert PDN, Kevan PG, Packer L (2009) DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Molecular Ecology Resources 9(s1): 196-207. https://doi.org/10.1111/j.1755-0998.2009.02645.x

Schindler M, Diestelhorst O, Härtel S, Saure C, Schanowski A, Schwenninger HR (2013) Monitoring agricultural ecosystems by using wild bees as environmental indicators. BioRisk 8: 53–71. https://doi.org/10.3897/biorisk.8.3600

Schmidt S, Schmid-Egger C, Morinière J, Haszprunar G, Hebert PDN (2015) DNA barcoding largely supports 250 years of classical taxonomy: Identifications for Central European bees (Hymenoptera, Apoidea partim). Molecular Ecology Resources 15: 985–1000. https://doi.org/10.1111/1755-0998.12363

Schneider N (2018) Recension douvrage: Découvrir & protéger nos abeilles sauvages. Bulletin de la Société des naturalistes luxembourgeois 120: 163–164.

Steffan-Dewenter I, Klein AM, Gaebele V, Alfert T, Tscharntke T (2006). Bee diversity and plant-pollinator interactions in fragmented landscapes. In: Waser NM, Ollerton J (Eds) Plant-Pollinator Interactions. From specialization to generalization. The University Of Chicago (Chicago): 387–410.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21: 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tang M, Hardman CJ, Ji Y, Meng G, Liu S, Tan M, Yang S, Moss ED, Wang J, Yang C, Bruce C, Nevard T, Potts SG, Zhou X, Yu DW (2015) High-throughput monitoring of wild bee diversity and abundance via mitogenomics. Methods in Ecology and Evolution 6: 1034–1043. https://doi.org/10.1111/2041-210X.12416

van der Loos LM, Nijland R (2021) Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology 30: 3270–3288. https://doi.org/10.1111/mec.15592

VDI-Richtlinie 4340-1 (2021) Standardisierte bestandsschonende Erfassung von Wildbienen für ein Langzeitmonitoring. Verein Deutscher Ingenieure e.V., Düsseldorf. https://www.vdi.de/richtlinien/details/vdi-4340-blatt-1-biodiversitaet-standardisierte-bestandsschonende-erfassung-von-wildbienen-fuer-ein-langzeitmonitoring

Vereecken N (2018) Découvrir & protéger nos abeilles sauvages. Glénat: 1–192.

Wagner DL (2020) Insect declines in the anthropocene. Annual Review of Entomology 65: 457–480. https://doi.org/10.1146/annurev-ento-011019-025151

Weigand A, Herrera-Mesías F (2020) First record of the wild bee Eucera (Tetralonia) alticincta (Lepeletier, 1841) in Luxembourg. Bulletin de la Société des naturalistes luxembourgeois 122: 141–146.

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra KDB, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, van der Hoorn BB, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher JN, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price BW, Rinkevich B, Teixeira MAL, Várbíró G, Ekrem T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of the Total Environment, 678 pp. https://doi.org/10.1016/j.scitotenv.2019.04.247

Weigand AM, Macher JN (2018) A DNA metabarcoding protocol for hyporheic freshwater meiofauna: Evaluating highly degenerate COI primers and replication strategy. Metabarcoding and Metagenomics 2: e26869. https://doi.org/10.3897/mbmg.2.26869

Westphal C, Bommarco R, Carré G, Lamborn E, Morison N, Petanidou T, Potts SG, Roberts SPM, Szentgyörgyi H, Tscheulin T, Vaissière BE, Woyciechowski M, Biesmeuer JC, Kunin WE, Settele J, Steffan-Dewenter I (2008) Measuring bee diversity in different European habitats and biogeographical regions. Ecological Monographs 78: 653–671. https://doi.org/10.1890/07-1292.1

Westrich P (2019) Die Wildbienen Deutschlands - 2., aktualisierte Auflage. Verlag Eugen Ulmer (Stuttgart), 1–824

Westrich P, Frommer U, Mandery K, Riemann H, Ruhnke H, Saure C, Voith J (2011): Rote Liste und Gesamtartenliste der Bienen (Hymenoptera: Apidae) Deutschlands. In: Binot-Hafke M, Balzer S, Becker N, Gruttke H, Haupt H, Hofbauer N, Ludwig G, Matzke-Hajek G, Strauch M (Eds.) Rote Liste gefährdeter Tiere, Pflanzen und Pilze Deutschlands, Band 3: Wirbellose Tiere (Teil 1). Naturschutz und Biologische Vielfalt. Landwirtschaftsverlag (Münster) 70: 373–416.

Winfree R, Aguilar R, Vázquez DP, Lebuhn G, Aizen MA (2009) A meta-analysis of bees' responses to anthropogenic disturbance. Ecology 90: 2068–2076. https://doi.org/10.1890/08-1245.1

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3: 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. GigaScience 2(1): 2047-217X-2-4. https://doi.org/10.1186/2047-217X-2-4

Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, De Barba M, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Kauserud H, Orlando L, Pansu J, Pawlowski J, Tedersoo L, Thomsen PF, Willerslev E, Taberlet P (2019) DNA metabarcoding – Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology, 28 pp. https://doi.org/10.1111/mec.15060

Supplementary material 1

Supplementary file 1

Authors: Fernanda Herrera-Mesías, Christopher Bause, Sophie Ogan, Hannah Burger, Manfred Ayasse, Alexander M. Weigand, Thomas Eltz

Data type: Docx file.

Explanation note: Curations made to merge taxonomic lists resulting from DNA and *PIN* (metabarcoding pipeline part).

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/jhr.93.86723.suppl1

Supplementary material 2

Supplementary file 2

Authors: Fernanda Herrera-Mesías, Christopher Bause, Sophie Ogan, Hannah Burger, Manfred Ayasse, Alexander M. Weigand, Thomas Eltz

Data type: Docx file.

Explanation note: Barcoding of two critical specimens for the consensus list (CON).

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/jhr.93.86723.suppl2

Supplementary material 3

Supplementary file 3

Authors: Fernanda Herrera-Mesías, Christopher Bause, Sophie Ogan, Hannah Burger, Manfred Ayasse, Alexander M. Weigand, Thomas Eltz

Data type: Xlsx file.

Explanation note: Raw Hymenoptera metabarcoding data with number of sequence reads as well as information on taxon curation and individual tagging combinations.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/jhr.93.86723.suppl3

Supplementary material 4

Supplementary file 4

Authors: Fernanda Herrera-Mesías, Christopher Bause, Sophie Ogan, Hannah Burger, Manfred Ayasse, Alexander M. Weigand, Thomas Eltz

Data type: Docx file.

Explanation note: Identification results of *DNA* and *PIN*.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/jhr.93.86723.suppl4

Supplementary material 5

Supplementary file 5
Authors: Fernanda Herrera-Mesías, Christopher Bause, Sophie Ogan, Hannah Burger, Manfred Ayasse, Alexander M. Weigand, Thomas Eltz
Data type: Docx file.
Explanation note: Identification results of *IVI* and *PIN* experts per bee specimen compared to the CONsensus list.

Link: https://doi.org/10.3897/jhr.93.86723.suppl5